

Osservazioni sulla lettura delle tavole di ripartizione e dei quantili.

Chi è nato con la macchina calcolatrice tascabile in mano generalmente non capisce perchè mai a scuola si sia obbligati a imparare a memoria le tabelline... “tanto c’è la calcolatrice!”.

Per la normale i dubbi sono i medesimi: perchè mai tutte queste formule quando con l’ausilio di un calcolatore è immediato ricavare $P(X \leq t) \quad X \sim N(\mu, \sigma^2)$?

La risposta è identica: prima dell’avvento dei calcolatori e ancora adesso quando il calcolatore non è disponibile (leggasi all’esame!) l’unico modo per arrivare ad un risultato numerico è quello di ricondurre la probabilità cercata alla funzione di ripartizione della normale standard i cui valori (non tutti ma solo quelli positivi...) sono tabulati in tavole che si trovano generalmente in fondo ai libri di testo.

Poichè $t, \mu \in \mathbb{R}$ e $\sigma^2 \in \mathbb{R}^+$ servirebbero ∞^3 tavole distinte!

Sfruttando le proprietà delle specifiche funzioni di ripartizione (o delle relative inverse) è possibile ottimizzare le tavole riducendone la dimensione complessiva ai soli casi essenziali.

Ad esempio non sarà necessario che le tavole della funzione di ripartizione gaussiana riportino $\Phi(z)$ per valori negativi di z in quanto questi saranno comunque ottenibili sfruttando la simmetria della distribuzione normale con la relazione fondamentale: $\Phi(z) = 1 - \Phi(-z)$.

Se questa proprietà da un lato ci consente di ridurre le dimensioni della tavola, dall’altro fornisce arbitrarietà nella scelta dei dati da riportare sulla tavola. Grazie ad essa, infatti, nulla vieta all’autore di riportare, ad esempio, $\Phi(z)$ per i valori negativi di z e di lasciare al lettore il calcolo per valori positivi sfruttando la relazione di simmetria.

E poichè non esistono convenzioni a riguardo, se non quelle dettate dalla logica, ecco che su alcuni testi troviamo $\Phi(z)$ tabulata per $z > 0$, su altri per $z < 0$ o, addirittura, $\Phi(z) - 0.5$ tabulata per $z > 0$.

Ecco perchè ogni tavola è dotata di una intestazione che specifica i dati esposti. Ciò può essere fatto o indicando l’espressione analitica della funzione tabulata o, più semplicemente, mettendo il grafico della funzione di densità in cui viene annerita la regione di probabilità di interesse.

Oltre alle tavole della funzione di ripartizione della normale standard, esistono anche le tavole dei quantili delle distribuzioni generalmente utilizzate in statistica inferenziale ovvero chi quadro $\chi^2(n)$, t di student $t(n)$, Fisher $F(n, m)$ e Kolmogorov-Smirnov $K(n)$ (a volte al posto di tabulare la distribuzione di Kolmogorov-Smirnov viene tabulata direttamente la statistica D_n che fornisce la soglia critica nel test omonimo).

Tavole della funzione di ripartizione

La distribuzione sempre presente è quella gaussiana ma in molti testi si trovano anche tavole per la funzione di probabilità o o la funzione di ripartizione binomiale e/o poissoniana. Noi ci occuperemo solo della gaussiana.

Gaussiana

Ecco come utilizzare l'unica tavola disponibile per ricavare il risultato numerico (approssimato) di qualunque valore di probabilità normale.

Vi avviso che farlo è molto più semplice che dirlo...

1° passo – riconduzione agli intervalli elementari

Detto I_X l'intervallo la cui probabilità dobbiamo misurare, potrà capitare che l'intervallo sia illimitato a destra $I_X \equiv (a, +\infty)$ o a sinistra $I_X \equiv (-\infty, b)$, limitato generico $I_X \equiv (a, b)$, limitato simmetrico rispetto alla media $I_X \equiv (\mu - a, \mu + a)$ o, addirittura, sia la composizione di alcuni dei precedenti.

Comunque sia, ci si può sempre ricondurre ai seguenti casi “elementari”:

- $(X \leq t)$ “coda di sinistra”
- $(X > t)$ “coda di destra”
- $(|X| \leq t)$ “regione centrale simmetrica intorno alla media”
- $(|X| > t)$ “code simmetriche”

Ad esempio $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$.

Notare che, essendo nel caso continuo, possiamo tranquillamente aggiungere o togliere il simbolo di uguale a piacimento senza alterare il risultato perchè se X è continua $P(X = a) = 0$.

2° passo - standardizzazione

A questo punto sfruttiamo la proprietà delle v.a. normali:

se $X \sim N(\mu, \sigma^2)$ e $Z \sim N(0,1)$ allora $X = \mu + \sigma Z$ o, equivalentemente, $Z = \frac{X - \mu}{\sigma}$.

e riconduciamo ognuno dei quattro casi al calcolo di una probabilità per una v.a. normale standard:

$$P(X \leq t) = P\left(Z \leq \frac{t - \mu}{\sigma}\right)$$

$$P(X > t) = P\left(Z > \frac{t - \mu}{\sigma}\right)$$

$$P(|X| \leq t) = P\left(|Z| \leq \frac{t - \mu}{\sigma}\right)$$

$$P(|X| > t) = P\left(|Z| > \frac{t - \mu}{\sigma}\right)$$

3° passo – riconduzione alla fdr

ora non ci resta che scrivere tutto in funzione della ripartizione normale standard $\Phi(z)$

dove $z = \frac{t - \mu}{\sigma}$:

$$P(Z \leq z) = \Phi(z)$$

$$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$$

$$P(|Z| \leq z) = P(Z \leq z) - P(Z \leq -z) = 2\Phi(z) - 1$$

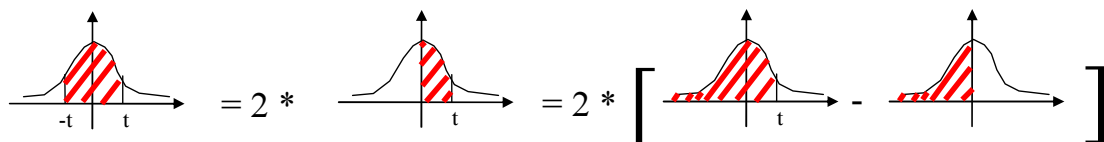
$$P(|Z| > z) = 1 - P(|Z| \leq z) = 2 - 2\Phi(z)$$

Ho esplicitato i passaggi intermedi perchè, secondo me, ricostruire all'occorrenza le formule è più facile che impararle a memoria.

La prima formula coincide con la definizione di ripartizione, la seconda e la quarta sono il complemento a 1 delle rispettive precedenti.

La terza formula è l'unica che merita un approfondimento grafico.

Abbiamo disponibile un valore t e vogliamo calcolare la probabilità sull'intervallo simmetrico $(-t, t)$ a partire dalla probabilità sull'intervallo $(-\infty, t)$ che sappiamo trovare tramite tavole. Ecco come ricostruire al volo la formula con un "disegno":



4° passo – adattamento alle tavole

adesso che ci siamo ricondotti a $\Phi(z)$ dobbiamo verificare il tipo di tavole a nostra disposizione.

Ogni tavola deve specificare nell'intestazione quale sia la funzione tabulata e, come abbiamo già sottolineato, non tutte le tavole tabulano proprio $\Phi(z) = \int_{-\infty}^z \varphi(t) dt$.

Qualche tavola riporta $1 - \Phi(z) = \int_z^{+\infty} \varphi(t) dt$ e addirittura a volte si trovano tavole che tabulano $\Phi(z) - \frac{1}{2} = \int_0^z \varphi(t) dt$.

Non solo. Quasi tutte le tavole riportano solo i valori per $z > 0$. Quindi, mi raccomando, prima di partire in quarta e sbagliare, controllate con che tavole avete a che fare.

Nell'ipotesi che le tavole a disposizione siano le più comuni ovvero riportino $\Phi(z) = \int_{-\infty}^z \varphi(t) dt$ per $z > 0$ ecco cosa dovremo fare:

se il valore in nostro possesso è $z > 0$ cerchiamo direttamente $\Phi(z)$ se, invece, $z < 0$ cerchiamo $\Phi(-z)$ e ricaviamo poi $\Phi(z)$ sfruttando la relazione di simmetria $\Phi(z) = 1 - \Phi(-z)$

5° (e ultimo) passo – ricerca nelle tavole

Supponiamo che, dopo tutti i passi precedenti, si sia arrivati a concludere che dobbiamo cercare un certo $\Phi(t)$.

Generalmente le tavole sono strutturate in tabelle con 2 o 11 colonne.

In entrambi i casi la prima colonna contiene i valori di t .

Se la tavola è a due colonne non ci resta che cercare nella prima colonna il valore che meglio approssima il nostro t e, in corrispondenza della stessa riga, leggere nella seconda colonna il valore $\Phi(t)$ cercato.

Se la tavola è a 11 colonne, la prima conterrà ancora i valori di t (presumibilmente troncati al primo decimale) mentre le dieci colonne successive conterranno i valori di $\Phi(t)$ con un ulteriore decimale di approssimazione.

Se ad esempio siamo sulla riga di $t = 1.2$ nella prima colonna leggeremo 1.2 mentre nelle colonne successive troveremo $\Phi(1.20), \Phi(1.21), \Phi(1.22), \dots, \Phi(1.29)$.

Se t espresso con due decimali non è sufficiente (perchè siete pignoli o perchè il prof. è “esigente”) potrete guadagnare un briciolo di precisione in più effettuando l'interpolazione lineare dei due valori che trovate sulle tavole vicini al vostro t .

In mancanza delle tavole dei quantili della gaussiana, questa tavola può essere usata a rovescio per il calcolo di un quantile. Si tratterà quindi di cercare il livello del quantile all'interno della tavola e individuare poi il relativo quantile nelle corrispondenti intestazioni di riga e di colonna.

Tavole dei quantili

Tutti i testi riportano le tavole dei quantili per le distribuzioni gaussiana, t di student e chi quadro.

Raramente i quantili della gaussiana sono omessi perchè individuabili tramite l'uso inverso della tavola della funzione di ripartizione gaussiana (vedi sopra).

Molti testi riportano anche i quantili di altre distribuzioni usate nei test di ipotesi come la Fisher o la Kolmogorov-Smirnov.

Gaussiana

Tutti gli autori riportano il quantile vero e proprio (cioè il valore z_α che lascia alla sua sinistra una probabilità pari ad α ovvero il valore z_α che rende $P(Z < z_\alpha) = \alpha$ un'identità).

La tavola è di facile lettura:

- nella prima riga cerchiamo i valori di α (in genere sono disponibili quelli più usati ovvero 0.8, 0.9, 0.95, 0.975, 0.99, 0.995)
- nella seconda riga troviamo i corrispondenti valori di z_α

Ancora vale la relazione fondamentale di simmetria $z_\alpha = -z_{1-\alpha}$ che ci serve nel caso dovessimo ricavare quantili per valori di $\alpha < 0.5$ (ad es. $z_{0.01} = -z_{0.99}$).

Inoltre possiamo sfruttare le altre relazioni di simmetria per ricavare il quantile quando α rappresenta la probabilità di un intervallo generico:

Prima si procede con la standardizzazione della relazione per arrivare a una delle quattro forme solite:

$$\alpha = P(Z \leq z) = \Phi(z)$$

$$\alpha = P(Z > z) = 1 - \Phi(z)$$

$$\alpha = P(|Z| \leq z) = 2\Phi(z) - 1$$

$$\alpha = P(|Z| > z) = 2 - 2\Phi(z)$$

dove però, questa volta l'incognita non sarà α ma z .

Esplicitando l'incognita ricaviamo le formule per il calcolo del quantile nei quattro casi:

$$\text{se } \alpha = P(Z \leq z) \Rightarrow z = z_\beta = \Phi^{-1}(\alpha) = z_\alpha$$

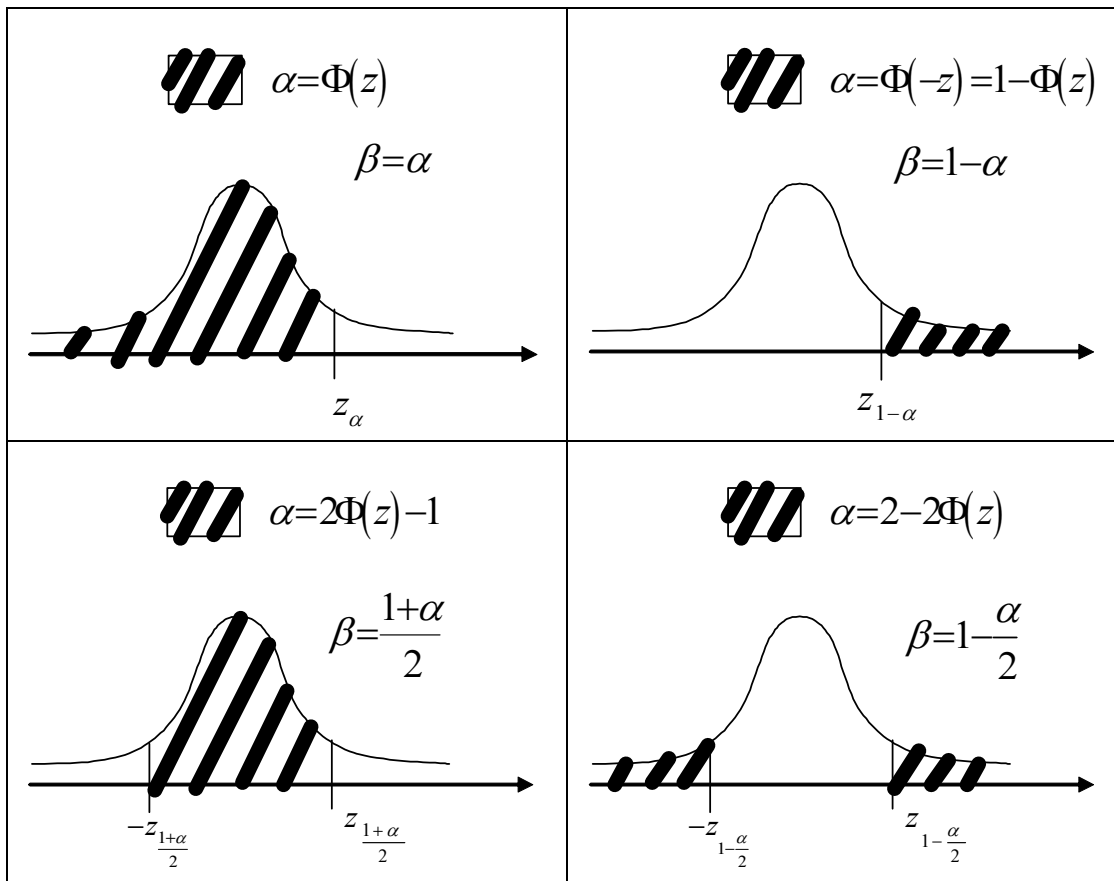
$$\text{se } \alpha = P(Z > z) \Rightarrow z = z_\beta = \Phi^{-1}(1 - \alpha) = z_{1-\alpha}$$

$$\text{se } \alpha = P(|Z| \leq z) \Rightarrow z = z_\beta = \Phi^{-1}\left(\frac{\alpha + 1}{2}\right) = z_{\frac{\alpha+1}{2}}$$

$$\text{se } \alpha = P(|Z| > z) \Rightarrow z = z_\beta = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = z_{1-\frac{\alpha}{2}}$$

Le precedenti formule, riferendomi ad esempio alla seconda delle quattro, vanno lette in questo modo:

Se α è il peso della coda di destra allora il valore z che lascia alla sua destra proprio una regione di peso α è il quantile z_β il cui livello β è determinato dal peso α tramite la relazione $\beta = 1 - \alpha$.



Alcuni autori sollevano il lettore dalle conversioni riportando sulle tavole i quantili direttamente per tutti e quattro i casi.

t di student

Poichè la distribuzione è simmetrica come la gaussiana, quanto detto sinora per la tavola dei quantili della gaussiana si applica anche alla tavola dei quantili della t di student.

In aggiunta abbiamo che la t di student dipende da un parametro (i gradi n di libertà) quindi sotto alla riga che riporta i principali valori di α troveremo una riga che riporta i valori dei quantili, indicati indifferentemente con $t_\alpha(n)$ o con $t_n(\alpha)$, per ogni valore di n (che leggiamo nella prima colonna).

Quando $n > 120$ la t di student può essere approssimata con una gaussiana quindi se dobbiamo calcolare un quantile per $n > 120$ e l'autore, gentilmente, ha riportato un'ultima riga con $n = \infty$ leggeremo il nostro quantile direttamente nell'ultima riga altrimenti ce lo andiamo a cercare nella tavola dei quantili della normale standard.

Qui è importante un'osservazione: molti autori non riportano il quantile vero e proprio. Tali autori indicano con $t_\alpha(n)$ il valore che lascia una probabilità α non alla sua *sinistra* ma alla sua *destra*.

In ogni caso l'intestazione della tavola indica chiaramente la scelta operata dall'autore. Si tratterà quindi di leggere attentamente ed effettuare le opportune conversioni già viste prima e che per comodità riportiamo anche qui di seguito.

Relazione di simmetria: $t_\alpha(n) = -t_{1-\alpha}(n)$.

Relazioni per ricavare il quantile quando α rappresenta la probabilità di un intervallo generico:

se $\alpha = P(t(n) < t) \Rightarrow t = t_\alpha(n)$	se $\alpha = P(t(n) > t) \Rightarrow t = t_{1-\alpha}(n)$
se $\alpha = P(t(n) < t) \Rightarrow t = t_{\frac{1+\alpha}{2}}(n)$	se $\alpha = P(t(n) > t) \Rightarrow t = t_{\frac{1-\alpha}{2}}(n)$

Chi quadro

Poichè anche la chi quadro, come la t di student, dipende dai gradi di libertà, le tavole dei quantili della chi quadro avranno il medesimo aspetto di quelle della t di student.

Anche in questo caso molti autori non riportano la coda sinistra ma quella di destra quindi sarà opportuno anche qui fare attenzione durante la lettura.

Inoltre, rispetto alle precedenti distribuzioni, la chi quadro non presenta simmetrie e dunque le relazioni di simmetria non sono valide.

Ecco perchè sulle tavole della chi quadro, accanto ai valori di probabilità più comuni trovano in genere posto anche i relativi valori complementari (ovvero per ogni valore α in genere si trova pure il corrispondente valore complementare $1 - \alpha$).

Fisher

Dipende da due parametri n, m (entrambi rappresentanti dei gradi di libertà) quindi la tavola è leggermente più complessa delle precedenti poichè, in questo caso, avremo una tavola per ogni valore α . Si tratta quindi di scegliere prima la tavola corrispondente al valore di probabilità α . Poi, all'interno della tavola scelta, cercheremo i gradi di libertà tra le intestazioni di riga e di colonna. All'incrocio troveremo il quantile cercato.

Anche questa distribuzione non è simmetrica quindi non è possibile usare relazioni di simmetria.

Esiste, però, una relazione che ci consente di passare dai quantili superiori (quelli con α prossimo a 1) a quelli inferiori (quelli con α prossimo a 0): $f_{1-\alpha}(n, m) = 1 / f_{\alpha}(m, n)$ (attenzione che oltre all'inversione abbiamo anche n e m che si scambiano di posto!).

Anche in questo caso molti autori non riportano la coda sinistra ma quella di destra quindi sarà opportuno anche qui fare attenzione durante la lettura.

Kolmogorov-Smirnov

La distribuzione dipende da un parametro e non è simmetrica. Valgono quindi tutte le considerazioni fatte per la tavola dei quantili della chi quadro.